Genome-wide Analyses of *Bacillus amyloliquefaciens* Strains Provide Insights into their Beneficial Role on Plants

Adnan Niazi

*Faculty of Veterinary Medicine and Animal Science*
*Department of Animal Breeding and Genetics*
*Uppsala*

Doctoral Thesis
Swedish University of Agricultural Sciences
Uppsala 2014
Genome-wide Analyses of *Bacillus amyloliquefaciens* strains Provide Insights into their Beneficial Role on Plants

Abstract

The innate immunity system of plants provides a basal defence barrier to most microorganisms. However, many plant pathogens have evolved to overcome this defence. Certain bacteria in the PGPB (plant growth promoting bacteria) category that improve plant growth have also been found to improve plant defence against insect pests and pathogens. Some bacteria of the genus *Bacillus* are known to be associated with plant roots, and have potential as possible biocontrol agents and biofertilizers in agriculture. For instance, *Bacillus amyloliquefaciens* subsp. *plantarum* strains can support plant growth and protection to stress after developing physical and biochemical contacts with plants. This thesis provides detailed descriptions of the genomic structure of three *B. amyloliquefaciens* subsp. *plantarum* strains with ability to promote plant growth and to suppress disease from several pathogens. The *Bacillus* genomes contain the basic genetic traits required for survival in the rhizosphere and plant growth promotion including chemotaxis and motility, root colonization, and biosynthesis of phytohormones. Besides growth promotion, the genomes have the capability to encode several antibacterial and antifungal compounds that effectively protect plant from pathogenic microorganisms. Several of the predicted traits were confirmed by experimental analysis.

Genome-wide comparative analysis of the *Bacillus* strains indicates that the genomes are very similar although variation has been observed in phenotypes associated with plant growth promotion and disease suppression. A possible explanation could be mutations in one or more putative genes. Genomic comparison with other non-plant associated *Bacillus* species indicates the genomic polymorphism has a crucial role in loss and gain of function in the two groups of *Bacillus* species.

Keywords: *Bacillus*, genome, plant-growth promotion, biocontrol, genetic diversity.

Author’s address: Adnan Niazi, SLU, Department of Animal Breeding and Genetics, P.O. Box 7023, 750 07 Uppsala, Sweden

E-mail: Adnan.Niazi@slu.se
Dedication

To my parents …

*The glory of science is to imagine more than we can prove.*

Freeman Dyson
## Contents

List of Publications 7

Abbreviations 9

1 Background 11

2 Introduction 13

2.1 Brassica oilseed crops 13

2.2 Arabidopsis as a model plant 13

2.3 Biocontrol 14

2.4 Plant-bacteria interactions and biocontrol 14

2.5 Induced Plant Resistance/ Priming of plant defence 15

2.6 Plant-associated *Bacillus* species – their potential effects and uses in agriculture 16

2.6.1 Direct growth promotion 17

2.6.2 Plant protection and sustainability 17

2.7 *Bacillus amyloliquefaciens* and plants 19

2.8 Bioinformatics and Next Generation Sequencing and Genomes 20

3 Objectives 21

4 Results and Discussion 23

4.1 Genome analysis 23

4.1.1 General characteristics of UCMB strains 23

4.1.2 Genomes of plant-associated UCMB isolates (Paper I – IV) 23

4.1.3 Antagonistic characteristics 24

4.1.4 Resistance against antibiotics 26

4.2 Genetic traits involved in plant association and growth promotion (Paper III & IV) 28

4.2.1 Chemotaxis and motility 28

4.2.2 Biofilm formation, root colonization and establishment 28

4.2.3 Phytohormones and Volatiles 29

4.2.4 Exo-enzymes – tools for host interaction and virulence 29

4.2.5 Type VII secretion system 30

4.3 Genomic diversity of *B. amyloliquefaciens* (Paper III & IV) 31

4.3.1 Genomic Comparison of the two subspecies/Inter-species comparison 31

4.3.2 Genetic polymorphisms 31
List of Publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:


IV Adnan Niazi, Sarosh Bejai, Khurram Maqbool, Shahid Manzoor, Johan Meijer, Erik Bongcam-Rudloff. Comparative genomics of *Bacillus amyloliquefaciens* species highlights potential determinants of plant association and interaction (manuscript)

Papers I-III are reproduced with the permission of the publishers. Papers I-III are reproduced with the permission of the publishers.
The contribution of Adnan Niazi to the papers included in this thesis was as follows:

I  Partly planned the study, contributed in genome assembly and annotation, and manuscript drafting.

II  Partly planned the study, assembled the genome and performed structural annotation, and wrote manuscript draft.

III Majorly planned the study, genome assembly and annotation, analysed genome and wrote the manuscript with comments and suggestions from co-authors.

IV Planned the study and conducted experiments, performed *in silico* comparative analysis and contributed towards the manuscript drafting with comments and suggestion from co-authors.
Abbreviations

AMP   Antimicrobial Peptide
BCA   Biological Control Agent
cLP   cyclic Lipopeptide
IAA   Indole-3-Acetic Acid
ISR   Induced Systemic Resistance
JA    Jasmonic Acid
MAMP  Microbe-Associated Molecular Patterns
NRPS  Non-Ribosomal Peptide Synthetase
PGPR  Plant Growth-Promoting Rhizobacteria
PK    Polyketide
PKS   Polyketide Synthetase
SA    Salicylic Acid
SAR   Systemic Acquired Resistance
SNP   Single Nucleotide Polymorphism
UCM   Ukrainian Collection of Microorganisms
VOC   Volatile Organic Compound
1 Background

Food crops in modern agriculture face many challenges with respect to productivity and security. Approximately 16% of the total crop yield is spoiled globally due to different diseases (Evans et al., 2010). Advancements in agricultural research have provided solutions like chemical fertilizers and pesticides, traditional breeding and genetic modification of plants, which have contributed significantly to plant quality and productivity. But economical unfeasibility and other drawbacks of application of these methods such as nutrient leakage, pesticide resistance in insects, and adverse effects on health and environment demand alternative approaches to overcome such problems and concurrently maintain food safety and security.

Many soil microorganisms have potential to increase soil fertility and plant health, but their roles have been disregarded for long time in intensive agriculture. Plant associated non-pathogenic microorganisms can be used to achieve sustainable increase in crop yields and to replace chemical fertilizers and pesticides. However, their successful application in agriculture as an alternative tool confronts many scientific challenges.

Recent development of genomic and high-throughput sequencing approaches of organisms allows exploration of genomic features and genetic traits involved in complex processes related to e.g. physiology, adaptation, interaction and association. The availability of genomic sequences of many microorganisms and economically important plants, has laid the foundation to understand key biological processes involved in plant growth promotion and antagonism against pathogens. This provides the opportunity to reveal the genetic traits involved in complex interactions between plants and microbes. Bacteria-mediated biocontrol is one such strategy that has great promise to cope with challenges faced by plants but the variation in efficacy and the underlying mechanisms among different bacterial species, is still poorly understood. This thesis provides a genomic overview of three Bacillus
*Pseudomonas amyloliqufaciens* isolates, and possible genetic traits underlying plant growth promotion and antagonism against pathogens.
2 Introduction

2.1 Brassica oilseed crops

Oilseed rape (*Brassica napus* L.), a species of the *Brassicaceae* family, is one of a main oilseed crops grown all over the world that accounts for approximately 13% of global oilseed production (FAO 2013). Another related less common oil crop is *Brassica rapa* which together with *Brassica oleracea* formed the amphidiploid species *B. napus*. Rapeseed oil is used for cooking oil and biofuel production while the press cake can be used as protein rich animal fodder. According to the United States Department of Agriculture (USDA), rapeseed was the third leading source of vegetable oil in the world in 2000, after soybean and palm oil, as well as the world’s second largest source of protein meal. Due to its worldwide consumption and utilization, the global demand of *Brassica* crops has increased rapidly. Unfortunately, the crop has become vulnerable to many serious pests and diseases that cause loss in yield and quality of *Brassica* crops due to the intensification in cultivation of Brassicas. Some pathogens that have been reported to cause yield losses in oilseed rape are fungi such as, *Alternaria brassicae*, *Rhizoctonia solani*, *Verticillium dahliae*, *Leptosphaeria maculans*, *Sclerotinia sclerotiorum* and *Botrytis cinerea*, as well as insect pests like Diamondback moth (*Plutella xylostella*), pollen beetles (*Meligethes aeneus*) and flea beetles (*Phyllotreta spp.*).

2.2 Arabidopsis as a model plant

*Arabidopsis thaliana* is one of the most well-known members of the *Arabidopsis* genus and is extensively used in research. It is a small dicotyledonous weed that is mostly found in temperate regions of the world (Alonso-Blanco and Koornneef, 2000). It belongs to the *Brassicaceae* family
and is a close relative to *Brassica* species such as *B. napus*. *A. thaliana* has become firmly established as a plant model organism since the 1980s with the advantages of having a small and simple genome, rapid life cycle, prolific seed production and the availability of numerous mutants. It was the first plant to be completely genome sequenced in the year 2000. The genome has five chromosomes and a total size of approximately 119,14 Mb with 27,416 genes annotated thus far (TAIR, 2010).

Arabidopsis offers a great advantage when used in *Brassica* research due to its high sequence similarity with *Brassica* species, and has become the focus of a genome project to understand the biology of a plant at the molecular level. The availability of various ecotypes also offers an opportunity to study the effects of natural variation for example in relation to PGPR.

### 2.3 Biocontrol

Biocontrol is the use of organisms like predatory insects, nematodes, bacteria and fungi in a specific ecological niche to protect certain organisms like crop plants from diseases and pests. Such organisms used to minimize the undesirable effects of detrimental organisms (pests and pathogens) on plants are referred to as Biological Control Agents (BCAs). There exists a vast range of methodologies by which biocontrol can be mediated, for instance; releasing predatory insects in the field (Wang *et al.*, 1999); introducing bacteria and fungi on the aerial surface of plants by spraying, which produce toxin in the gut when taken up by pests and kills them (Roh *et al.*, 2007); and also by inoculating seeds with bacteria that allows them to colonize the rhizosphere and consequently protect plants from pathogens. The manipulation of microbial communities associated to plants has received attention in agriculture with the purpose to protect plants from diseases and increase crop yields (Banerjee *et al.*, 2006; Adesemoye *et al.*, 2008).

### 2.4 Plant-bacteria interactions and biocontrol

Survival of many bacteria is dependent on other organisms such as plants, by developing close relationships with them. Plant-microbe interactions are extremely diverse and generally poorly understood. Plants recognise microbe-derived molecules called as microbe associated molecular patterns (MAMPs) via the pattern recognition receptors (PRRs) followed by the activation of immune response. During the early phase of colonization of beneficial bacteria, the suppression of MAMP responses is critical to protect the plant by MAMP derived antimicrobial compounds (Millet *et al.*, 2007). Many soil bacteria are
heterotrophic that rely on external source of organic compounds available in soil. Dead animals and plant materials as well as compounds released by plant roots (root exudates) constitute many important carbon compounds such as carbohydrates, organic acids and amino acids (Lugtenberg et al., 2001). These carbon compounds are key contributions to soil resources that collectively support diverse bacterial communities. The root exudates not only constitute nutrient substrates for bacteria but also contain molecules that affect bacterial gene expression and behaviour giving the plants possibilities to shape the local soil microbiota. Secondary metabolites may antagonize many microbes keeping them away from the root tissues to avoid detrimental effects. Other signalling molecules in exudates are detected by certain bacteria using receptors that subsequently elicit signal-transduction mechanisms, which consequently mediate their response to the local environment (Brencic & Winans, 2005). The diffusion of signalling molecules in root exudates is restricted to a limited region of soil surrounding the root known as the rhizosphere. Bacteria that are present in the rhizosphere are called rhizobacteria. Certain rhizobacteria become enriched in the rhizosphere and develop close physical and biochemical contacts with the plant root surface, the rhizoplane. The primary source of nutrients for such bacteria is root exudates and in return the bacteria may support plants by stimulating growth and improving stress management (Yang et al., 2009).

Certain rhizobacteria help plants to acquire nutrients from soil and provide protection against biotic and abiotic stress, whereas pathogens utilize plant tissues and nutrients affecting growth and production of the host plants. Rhizobacteria that support plant growth are termed as Plant-Growth-Promoting Rhizobacteria (PGPR). Plant growth promotion by PGPR is mediated in several ways, for instance, by making adequate amount of macronutrients and micronutrients available for plants; and by producing growth stimulating hormones (Varma et al., 2004). Many PGPR have also been noted to confer plant protection against pests and pathogens (Zehnder et al., 2001; Conrath et al., 2006; Van Oosten et al., 2008). PGPR confer plant protection by depleting resources essential for the survival of microorganisms by synthesizing antibiotics and different enzymes such as, proteases and chitinases that compromise harmful pests and microorganisms, and by elicitation of defence responses (Whipps, 2001; Lugtenberg & Kamilova, 2009).

2.5 Induced Plant Resistance/ Priming of plant defence

Together with constitutive defence, plants have also evolved inducible defence mechanisms to overcome detrimental pathogens and tolerate abiotic stress.
Induced defence mechanisms occur in two major forms; Systemic Acquired Resistance (SAR) and Induced Systemic Resistance (ISR), as shown in Figure 1. Non-pathogenic rhizobacteria mediated ISR is similar to pathogen induced SAR where in the uninfected distal parts of plants are primed for enhanced resistance to pathogens. The mechanisms of induction of ISR and SAR are via different signalling pathways, though. Induction of SAR is through salicylic acid (SA) via the elicitation of PR1 genes while ISR requires jasmonic acid (JA) and ethylene (ET) (Van Loon et al., 1998; Pieterse et al., 2001). Also other compounds than the classical phytohormones may be involved in the stimulation of ISR. An example of this is the plant growth promoting Bacillus species producing 2,3-butanediol that elicits ISR in A. thaliana and eventually leads to a significantly lower level of disease incidence by pathogenic Erwinia carotova in Arabidopsis (Ryu et al., 2004). The enhanced ability of stimulating stronger resistance responses associated with ISR is called priming and is achieved by treatment of plants with certain chemicals of microorganisms (Conrath et al. 2002). PGPR primed plants exhibit enhanced defence responses are not costly but rather display an accelerated response to pathogen/pest invasion (Van Hulten et al., 2006; Pozo et al., 2008; Stein et al., 2008).

2.6 Plant-associated Bacillus species – their potential effects and uses in agriculture

Bacillus is a genus of gram positive and rod shaped bacteria. They are capable to form stable dormant structures called endospores in nutrient void and stressful environmental conditions. Spores are generally viable for a long period even under harsh conditions. Several PGPR based products are commercially available and many of them contain Bacillus strains (Kloepper et al., 2004). Bacillus is one of the dominant genera isolated from soil samples (Hallmann et al., 1999) along with Pseudomonas and Clostridium strains. The sporulation ability and easy cultivation of Bacillus species (Ross et al., 2001; Tiago et al., 2004) are attractive for their practical use as inoculants.

Many Bacillus species have been reported to exhibit positive effects on various plant species, such as oilseed rape (Danielsson et al., 2007; Bejai et al., 2009), wheat (El-Daim et al., 2014; Alvarez et al., 2012), tomato (Choudhary & Johri, 2009; Lim & Kim, 2009), and maize (Oliveira et al., 2009), mainly by suppressing pathogens. Among Bacillus, species like B. cereus, B. subtilis and B. amyloliquefaciens are known to support plant growth and protection, and has therefore gained attention as PGPR species.
2.6.1 Direct growth promotion

PGPR may operate through several mechanisms that directly effect plant growth. They can enhance plant growth directly by providing nutrients such as nitrogen, phosphate, and potassium to plants through nitrogen fixation or solubilising phosphorus (Kiers et al., 2003; Berg, 2009; Richardson et al., 2009). Also, many rhizobacteria are able to supply iron, by carrying out solubilisation to avoid precipitation of soil phosphorous (Gyaneshwar et al., 2002), and vitamins to plants (Richardson et al., 2009). In addition, several phytohormones such as auxins, ethylene, cytokinins and gibberellins are known to directly regulate physiological processes and enhance growth of the plant (Loper & Schroth, 1986; Varma et al., 2004; Berg, 2009). Plant growth-promoting Bacillus and other species are capable to biosynthesize such phytohormones or stimulate plant phytohormone synthesis. Indole-3-acetic acid (IAA), an auxin, produced by bacteria such as Azospirillum and Pseudomonas, is directly involved in enhancing plant growth (Loper & Schrot, 1986; Spaepen et al., 2007). Besides production of plant growth hormones, certain PGPR are able to synthesize volatile organic compounds (VOCs) such as acetoin and 2,3-butanediol that can trigger plant growth (Ryu et al., 2003; Ryu et al., 2004).

2.6.2 Plant protection and sustainability

Many PGPR support plant protection by preventing or attenuating negative effects due to abiotic and biotic stresses i.e. harmful effects caused by unfavourable environmental changes or phytopathogens. In this way PGPR make plants sustainable by avoiding costly energy investments into defence and repair of damaged tissues. The protective effects to pathogens can be achieved by plant-associated bacteria through antagonism against plant pathogens, competition in the rhizosphere for nutrients and space, and induction of ISR in plants. This is termed as biocontrol-mediated plant growth.

Antagonism

Antagonism involves suppression of plant pathogens through production of antimicrobial compounds against other microorganisms competing for nutrients and space. A wide range of antibiotics produced by different PGPR have been identified and characterized. Members of the Bacillus genus produce several phytopathogen-inhibiting compounds, such as kanosamine and zwittemycine A, surfactin, iturin, and fengycin, but proteases also play a determining role for competition for space and nutrients (Emmert & Handelsman, 1999; Ongena & Jacques, 2008; Hoeffer et al., 2012). Besides antibiotics, the antagonistic ability of bacteria may rely on exoenzymes such as
chitinases, having the capacity to degrade cell walls of fungal pathogens (Ordentlich et al., 1988; Inbar & Chet, 1991).

Figure 1. Schematic representation of induced immune responses in plants. Pathogen infection on local leaves leads to activation of systemic acquired resistance (SAR) in the systemic (non-infected) leaves. SAR is typically activated by the induction of SA leading to accumulation of pathogenesis related (PR) proteins with antimicrobial properties. Root colonising beneficial bacteria typically activate Induced systemic resistance (ISR). ISR expressing plants are commonly primed for accelerated jasmonic acid (JA) and ethylene (ET) signalling pathways. Both SAR and ISR expressing plants show enhanced defense capacity against a broad-spectrum of pathogens.

Competence
The rhizosphere is a reservoir of organic and inorganic bound carbon sources that are essential for bacterial survival and proliferation. The availability of energy rich carbon sources in the rhizosphere attracts beneficial as well as pathogenic microorganisms towards this environment, and they compete with each other for their survival. Antagonism is one of the strategies used by
rhizobacteria to kill or restrain other soil-borne organisms including pathogens to occupy growth space and thereby get access to essential nutrients. Besides antagonism activity, beneficial bacteria limit the growth of microorganisms by making iron unavailable through their efficient iron uptake systems. One of such mechanisms is the production of siderophores, iron-chelating compounds, which allow the producing bacteria to scavenge and solubilize iron complexes and subsequently deplete this micronutrient from the environment, thus suppressing growth of deleterious bacteria (Whipps, 2001). Many PGPR including *Bacillus* species are also able synthesize siderophores (Arguelles-Arias et al., 2009).

**2.7 Bacillus amyloliquefaciens and plants**

*B. amyloliquefaciens* species have been shown to efficiently colonize roots and overcome the antibacterial action of some plant root exudates (Reva et al. 2004). Various *B. amyloliquefaciens* isolates have also been shown to promote salt tolerance (Bochow et al. 2001), increase the yield of several plant species such as tomato (Guel et al. 2008), and enhance the root/shoot biomass of e.g. oilseed rape, wheat, and maize seedlings (Idriss et al. 2002; Asari et al., unpublished; El-Daim et al., 2014; Talboys et al., 2014). Certain strains of *B. amyloliquefaciens* such as FZB42, a type strain, synthesize secondary metabolites including cyclic lipopeptides (cLPs) and polyketides (PKs) with antimicrobial action (Koumoutsi et al., 2004; Chen et al., 2006; Ongena & Jacques, 2008). Such peptides may facilitate root colonization and interaction with host plant and prime plant defence responses (Ongena & Jacques, 2008).

Three *B. amyloliquefaciens* subsp. *plantarum* UCMB strains have been identified and characterized as PGPR. The plant growth-promoting efficacy of these strains has been investigated under different abiotic and biotic stress conditions. The results from different experiments revealed that all three bacterial strains had a significant effect on seed germination and seedling establishment of *Brassica napus* cv. Westar, wheat, and *A. thaliana* ecotype Col-0 (Asari et al., unpublished; El-Daim et al., 2014). The variability in plant growth promotion and pathogen inhibition was observed where UCMB5113 enhanced plant growth to a greater extent than the other two strains (Reva et al., 2004; Danielsson et al., 2007).
2.8 Bioinformatics and Next Generation Sequencing and Genomes

The recent progress in next generation sequencing (NGS) technologies such as Illumina (Bennett, 2004), 454 pyrosequencing (Margulies et al., 2005), IonTorrent (Rusk, 2011) and PacBio (Eid et al., 2009) has overcome the shortcomings of traditional Sanger sequencing by increasing throughput, scalability, speed, and resolution. This development has ignited a revolution in biological and biomedical research by increasing opportunities to intensively explore genomes of various organisms. Thousands of whole genome sequences of all three domains of life collectively have been deposited in public databases and more are becoming available. These data have significantly contributed to an improved understanding of the physiology, evolution and ecology of those organisms. Genome sequences of many PGPR isolates from the genera Bacillus, Pseudomonas, and Enterobacter are now available (Chen et al., 2007; Taghavi et al., 2009; Loper et al., 2012). This information can contribute to increase the knowledge on factors and mechanisms involved in biocontrol and plant-growth promotion. A great challenge is now to uncover and elucidate the genetic mechanisms involved in the plant-associated lifestyle and whole biocontrol process achieved by PGPR.

This thesis presents the genomes of plant-growth promoting rhizobacteria B. amyloliquefaciens UCMB isolates sequenced using Illumina and Ion Torrent NGS technologies, and genetic traits underlying biocontrol and plant growth enhancing activity.
3 Objectives

Plant growth promotion by rhizobacteria is a widely accepted phenomenon but the mode of action and determinants involved in functional mechanisms are not fully understood. The overall aim of the studies described in this thesis was to explain the genetic architecture of three \textit{B. amyloliquefaciens} isolates and to elucidate genetic traits that could be involved in stress management and plant growth promotion. It was also necessary to do a comparative study of the genomes with other \textit{Bacillus} species with known biocontrol ability and characterize their overall genomic differences.

The detailed objectives were:

- To characterize the genomes of three plant-beneficial \textit{B. amyloliquefaciens} UCMB isolates (Paper I – III).
- To investigate the importance of genetic diversity and variability of the \textit{B. amyloliquefaciens} strains in relation to plant growth promotion (Paper IV).
4 Results and Discussion

4.1 Genome analysis

4.1.1 General characteristics of UCMB strains

The three strains UCMB-5113, 5036, and 5033 are Gram-positive rod shaped, motile, spore forming, aerobic, and mesophilic bacteria. They are approximately 0.8 µm wide and 2 µm long and can grow on Luria Broth (LB) and potato dextrose agar (PDA) between 20 °C and 37 °C (28 °C optimal) within the pH range 4–8. These characteristics are probably required to survive under diverse environmental conditions. The bacterial strains were initially identified as members of the *B. amyloliquefaciens* group based on phenotypic analysis (Reva et al., 2004). *B. amyloliquefaciens* species were later proposed to be classified either as *B. amyloliquefaciens* subsp. *plantarum* (plant-associated) or as *B. amyloliquefaciens* subsp. *amyloliquefaciens* (non plant-associated) species (Borriss et al., 2011). The identity of the three strains was confirmed as *B. amyloliquefaciens* subsp. *plantarum* (Paper I–III). The phylogenetic tree where two sister groups, subsp. *plantarum* and subsp. *amyloliquefaciens* clustered separately, highlights the close relationship of UCMB strains with the *plantarum* family whereas subsp. *amyloliquefaciens* displayed less taxonomic proximity (Figure 2). UCMB5113 forms dark orange colonies while 5036 and 5033 lack the pigment and form pale yellow to white colonies. Like other rhizospheric and plant beneficial bacteria all three UCMB strains possess the ability to catabolize plant-derived compounds, colonize roots, and tolerate several heavy metals and antibiotics (Paper III).

4.1.2 Genomes of plant-associated UCMB isolates (Paper I – IV)

The genome sequences of three UCMB isolates 5113, 5036, and 5033 were assembled and annotated for structural and functional features. The chromosomal sizes of these bacteria are about 4.0 Mb. The number of protein...
coding genes in the UCMB5113 strain is less than the other two strains where UCMB5033 has more coding genes compared to the other two strains. Protein-coding genes were assigned functional categories based on clusters of orthologous group (COG). The percentage of genes in COG categories for carbon, energy, amino acid and nucleic acid metabolism were found to be high and similar in all three UCMB strains (Figure 3), indicating their efficient carbon, amino acid and nucleotide metabolism. This is an indication of their ability to utilise a wide range of nutrients that enables them to survive in the rhizosphere.

4.1.3 Antagonistic characteristics

The reduced growth of several pathogens in the presence of UCMB isolates under *in vitro* conditions previously demonstrated (Danielsson *et al.*, 2007) suggested the production of unknown bioactive compounds. Genome sequence availability reveals the capacity of the *Bacillus* strains to produce a high diversity of antifungal and antibacterial compounds. Several gene clusters directing non-ribosomal synthesis of secondary metabolites, predominantly cLPs and PKs, were found while other antimicrobial compounds are ribosomally synthesized. The potential to produce numerous antimicrobial peptides (AMPs) indicates a competitive advantage over other microbes and probably also support the development of a symbiotic relationship with plants (Ongena *et al.*, 2007; Collemare & Lebrun, 2012).

*Non-ribosomally synthesized bioactive metabolites*

*B. amyloliquefaciens* strains are known to synthesize cLPs and PKs by mechanisms not involving the ribosomal machinery otherwise used for peptide synthesis (Schneider *et al.*, 2007). The cLPs and PKs consist of monomeric building blocks as amino acids or organic acids that are linked together by modularly organized non-ribosomal peptide synthetases (NRPS) and polyketide synthases (PKS), respectively. Similarly, UCMB strains harbour genes that can synthesize non-ribosomal cLPs such as surfactin, fengycin, the iturin bacillomycin D (5113 & 5036) and iturin A (5033). These compounds are amphipathic molecules and vary in their peptide and fatty acid moieties (Hofemeister *et al.*, 2004). Surfactins are antibiotic compounds but they also possess hemolytic and surfactant properties (Kowall *et al.*, 1998), whereas fengycins are potent antifungal agents (Vanittanakom *et al.*, 1986) and iturin is a family of AMP with wide range of antibacterial and antifungal properties (Peypoux *et al.*, 1999). Besides cLPs, gene clusters for the production of three PKs: difficidin, macrolactin, and bacillaene are present in these strains. These PKs are known for their diverse antibacterial action (Romero-Tabarez *et al.*, 2007).
Other than cLPs and PKs, a cluster responsible for biosynthesis of bacilysin, a bioactive dipeptide, also exists. The synthesis of difficidin, bacillaene, and bacilysin AMPs by BCA are effective against the fire blight disease (Chen et al., 2009b). Except UCMB5036, both strains 5113 and 5033 possess an extra NRPS/PKS hybrid operon of unknown function in their genomes. Variability in the antagonistic activity against diverse phytopathogens was observed between the three strains where the inhibitory effect of UCMB5036 was higher in comparison with the other two isolates (Danielsson et al., 2007; Asari et al., unpublished). This could be due to the variation in the genetic structure in some of the operons due to deletion of genes, intra-species variation and differences in the length, which can alter configuration of the amino acids affecting the peptide moieties.

**Ribosomally synthesized antibiotics – Bacteriocins and Lantibiotics**

Bacteriocins are ribosomally synthesized proteinaceous toxins produced by several bacteria to inhibit the growth of similar or closely related bacterial strains. They are generally narrow spectrum antibiotic compounds. A circular bacteriocin, amylocyclicin, has antagonistic effect against closely related Gram-positive bacteria (Scholz 2014). The operon responsible for amylocyclicin production exists in the genomes of three UCMB strains analyzed. Another putative bacteriocin related gene, exclusively present in the UCMB5113 strain, has a DNase domain (PF12639) similar to those found in Colicin/pyocin S1/S2 toxins. A cannibalistic like behaviour was observed quite often for UCMB5113. This might be due to a three-gene operon similar to Sdp in *B. subtilis*, which probably encodes SdpC-like toxin (killing factor). SdpC toxin induces the lysis of sibling cells that have not entered the sporulation pathway, providing a source of nutrients to support sporulation (Ellermeier et al., 2006). UCMB5033 also has similar genes in its genome while 5036 lack this operon.

Lantibiotic (lanthionine-containing antibiotic) is a type of bacteriocins that form a particular group among the AMPs and are characterized by unique structural features. They are heat stable peptides produced by and act against Gram-positive bacteria by pore formation and/or inhibition of cell wall biosynthesis (Chatterjee et al., 1992; Sahl and Bierbaum, 1998). Mersacidin and amylolysin are the only lantibiotic AMPs known to be produced so far only by YAU B9601-Y2 and GA1 strains of *B. amyloliquefaciens* (He et al., 2012; Arias et al., 2013). The UCMB isolates analysed are impaired to produce these lantibiotics due to absence of synthesizing genes. However, UCMB5113 contains a gene cluster that probably encodes a novel lantibiotic peptide of unknown function (IV).
Figure 2. Phylogenetic tree highlighting the position of *B. amyloliquefaciens* isolates UCMB-5113, 5036 and 5033. The tree was based on core genome coding sequences (amino acids) and was constructed with Neighbour Joining method. *B. amyloliquefaciens* strains are grouped in single clade with two sub-clades, each for plant associated strains (green) and non plant-associated strains (pink).

4.1.4 Resistance against antibiotics

Besides the production of antibiotics, the three *B. amyloliquefaciens* strains concurrently protect themselves from antibiotic and toxic compounds produced by them or other microorganisms in the rhizosphere, which improves
competitiveness in microbial antagonism. The \textit{tetB} gene provides resistance to \textit{B. subtilis} against the antibiotic tetracycline by decreasing its accumulation in the cell (Sakaguchi \textit{et al.}, 1988). This gene is present in the genomes of the three strains and the function was confirmed by the good growth of UCMB5113 observed on tetracycline containing medium. Additionally, the \textit{sat}-4 gene that encodes streptothricin acetyltransferase in \textit{Campylobacter coli} BE/G4 is responsible for providing resistance against streptothricin (Jacob \textit{et al.}, 1994). The presence of the gene \textit{yyaR}, similar to \textit{sat}-4, probably implicates a role in streptothricin resistance. Bacitracin is a broad-spectrum antibiotic that is non-ribosomally synthesized by \textit{B. licheniformis} and \textit{B. subtilis} (Azevedo \textit{et al.}, 1993; Ishihara \textit{et al.}, 2002). The UCMB strains provide resistance via bacitracin-specific ABC transporter BceAB probably regulated by a two-component system, BceRS, located upstream of the operon.

![Figure 3](image-url)

**Figure 3.** COG classification of protein-coding genes of UCMB strains. The percentage of protein-coding genes classified in at least one COG group is labeled in the center. (A) 80.2% (2944/3656) coding sequences of the \textit{B. amyloliquefaciens} UCMB5113 genome were functionally classified in COG groups. (B) 81.1% (2945/3636) CDSs of the UCMB5036. (C) 77.1% (2984/3877) CDSs of the UCMB5033 genome.

Remarkably, the biosynthetic genes responsible for the production of any characterized lantibiotics are absent in the genomes of the three UCMB strains, but the traces of their immunity genes and regulators are present in their genomes. The two-component systems Mrsk2/MrsR2 or SpaK/SpaR induce expression of the immunity genes \textit{MrsEFG} or \textit{SpaEFG} upon detection of
mersacidin or subtilin produced by other competing microorganisms in the environment. Subtilin related immunity genes are present in all three UCMB strains, while genes providing protection from mersacidin are only present in UCMB5036.

4.2 Genetic traits involved in plant association and growth promotion (Paper III & IV)

4.2.1 Chemotaxis and motility
PGPR respond to changes in chemical composition of the environment, mainly in the concentration of plant-derived carbon compounds, by modulating the flagellar machinery to move towards the appropriate direction. The response is based on two-component signal transduction system that induces both genes for chemoreceptors (methyl-accepting chemotaxis proteins) and genetic determinants (che) that control flagellar rotation. All of the UCMB strains exhibit swarming motility due to presence of genes for intact pathways of chemotaxis (cheD, cheC, cheW, cheY, cheV, cheB, cheR), flagellar biosynthesis and components assembly (flg, flh, fli and mot operons) and regulatory gene swrA that up-regulates the expression of flagellar genes and increases swarming motility (Kearns et al., 2004; Ghelardi et al., 2012). The production of surfactin that lowers the surface tension and facilitates swarming motility is indispensable for swarming phenotype (Ghelardi et al., 2012). UCMB5113 has been observed to produce higher levels of surfactins as compared to the other two strains in drop collapse tests. These traits greatly favour host plant sensing and movement towards a suitable environment that enables proliferation of the UCMB strains.

4.2.2 Biofilm formation, root colonization and establishment
Adhesion to plant roots is a prerequisite for colonization and establishing associations with host plants. Bacterial adhesion on root surface may occur through factors like flagellin, externalized polysaccharides and adhesins leading to surface colonization. The UCMB strains, especially 5113, have been shown to colonize plant seeds and seedlings (Reva et al., 2004). This is indication of the presence of genes encoding proteins involved in surface adhesion. Genes present in the genomes of UCMB isolates putatively encode proteins with a collagen-like GXT structural motif, which is suggested to form fiber-like protein structures for adhesive purposes (Chen et al., 2007).

Biofilm formation is essential for efficient surface colonization. It is an extracellular matrix comprised of a variety of polysaccharides. UCMB isolates contains genes important for formation and regulation of biofilm process. For
instance, epsA-O and yqxM-sipW-tasA, which are required for the formation of biofilms in *B. subtilis* (Branda et al., 2006; Romero et al., 2011; Ostrowski et al., 2011). Biofilms have permeable water channels that allow exchange of nutrients and toxins and protection from different environmental stresses (Costerton et al., 1999; Davey & O'toole, 2000). It is also possible that biofilm produced by UCMB strains with a hydrophobic layer on the surface probably provide a protective shield to protect them as well as the host plant from pathogenic toxins and antimicrobial compounds.

4.2.3 Phytohormones and Volatiles

PGPR can influence plant growth directly by producing plant-growth promoting substances ‘phytohormones’, and their regulators, that modulate metabolism and plant physiology. Studies suggest that phytohormones like auxin and cytokinin produced by rhizobacteria have vital roles in the enhancement of plant growth and health (Salamone et al., 2001; Spaepen et al., 2007). The results presented in paper IV described the presence of genes encoding the biosynthesis of an auxin family compound, indole-3-acetic acid (IAA). Besides the production of phytohormones, these bacterial strains have been observed to directly influence plant growth by signalling (unpublished data). Volatile organic compounds (VOCs) are active organic compounds with high vapour pressure that play an important role in communication with plants (Farag et al., 2013). These strains possess genes encoding for different volatiles such as acetoin and 2,3-butanediol that not only directly enhance plant growth through airborne signals but also elicits induced systemic resistance (ISR) in plant (Ryu et al., 2003; Farag et al., 2013). It is possible that bacterial VOCs are among those factors, which initiate cross talks between bacteria and plants. Indirect plant-growth promotion is also possible through production of diverse compounds such as antibiotics, exo-enzymes, siderophores and lipopolysaccharides that suppress the growth of phytopathogens and degrades toxic compounds. Genetic traits required for biosynthesis of these compounds are conserved in UCMB isolates.

4.2.4 Exo-enzymes – tools for host interaction and virulence

Plant-associated bacteria secrete various enzymes to hydrolyze polysaccharides and other compounds available in the rhizosphere. Some exo-enzymes are featured with modules such as; glycoside hydrolases, glycosyltransferases, polysaccharide lyases, and carbohydrate esterases that catalyze the breakdown or modification of carbohydrates. These enzymes play important roles in acquiring nutrients as well as in interaction with bacteria, fungi and plants. The three UCMB strains contain such genes encoding enzymes for chitinase-like
activity and members of the polysaccharide lyase family such as pectate lyase (pel) and pectin lyase (pelB) facilitating degradation of the plant cell wall (Marin-Rodriguez et al., 2002). Such genes play important roles in host interaction as well as virulence but may also be used to process plant surfaces for establishment and colonization of plant roots by beneficial bacteria. UCMB5113 degrading activity on substrates such as amylase, phosphate, and urea compounds was positive. Besides these corresponding exo-enzymes, the genomes of the three UCMB strains encode proteins with Lysin motif (LysM). The LysM modules containing proteins in fungal pathogen Cladosporium fulvum have been suggested as potential “LysM effectors” that dampen host plant immunity (de Jonge & Thomma, 2009; de Jonge et al., 2010). UCMB5113 has three proteins, CwlS, lytF and an unknown protein, possessing LysM domains at the N-terminus and NlpC/60 at the C-terminal region. CwlS, an endopeptidase has been shown to be dependent on the number of LysM domains thus regulating efficient binding to PGNs (Wong et al., 2014). Expression of these genes in the presence of Arabidopsis and V. dahliae were examined to analyse plant-associated traits. The results revealed that in addition to the known function of LysM domains in PGN binding, proteins possessing these domains also attenuate the defence responses supporting an efficient colonization. Further studies are needed to verify their biological role in interactions with pathogens and host plants.

4.2.5 Type VII secretion system

Many proteins are translocated across the inner and outer membrane of cells using secretion or translocation systems. Type III and IV systems are generally found in Gram-negative bacteria that deliver virulence factors, “effectors”, into host organism. B. amyloliquefaciens species lack these secretion systems in their genomes but they indeed secrete proteins through the Sec pathway. The genomes of UCMB strains also possess an Esat-6-like secretion system (ESS) also known as the Type VII secretion system (T7SS). The role of T7SS in M. tuberculosis, S. aureus and B. anthracis is known in terms of bacterial pathogenicity (Guinn et al., 2004; Burts et al., 2005; Abdallah et al., 2007; Garufi et al., 2008). Secretion targets of this system are small proteins containing a WXG motif. T7SS-like apparatus in B. subtilis shares structural homology with S. aureus and was recently shown to mediate YukE (WxG motif) protein secretion (Huppert et al., 2014). The genomes of B. amyloliquefaciens subspecies possess an array of genes similar to T7SS. Expression of yukE is observed in UCMB5113 during exponential and stationary phases of growth, suggesting the system to be functional in this strain (Paper IV). But the biological role of T7SS in these strains is not known.
yet. Studying the role of this export pathway in the future might contribute to the understanding of new cellular roles.

4.3 Genomic diversity of B. amyloliquefaciens (Paper III & IV)

4.3.1 Genomic Comparison of the two subspecies/Inter-species comparison

The comparison of the genomes of subsp. plantarum and subsp. amyloliquefaciens strains shows that 72-84% of the coding genes of each strain are shared among the genomes. A high level of genomic homogeneity between the strains of plantarum and amyloliquefaciens indicates that these two groups have recently diverged. It is intriguing that amyloliquefaciens strains sequenced so far lack two large gene clusters in their genomes, encoding difficidin and macrolactin. Like cLPs, e.g. surfactins and fengycins (Ongena et al., 2007), conservation of PK-synthesizing operons in the plantarum group suggests a role in both virulence and avirulence activities. Genes unique to plant-associated strains probably distinguish bio-control strains from non plant-associated species. These include putative antibiotics, hydrolysing enzymes, transporters and proteins of unknown function (Figure 4), which may be important for the symbiotic lifestyle of the plantarum species.

![Graph showing product types of Plantarum specific protein-encoding genes.](image)

Figure 4. Product types of Plantarum specific protein-encoding genes.

4.3.2 Genetic polymorphisms

Single nucleotide polymorphisms (SNPs) appear to be key factors in understanding the adaptive process resulting in a lifestyle that supports plant association and protection. The genomic data revealed that variation between strains of the two subspecies was mainly due to the accumulation of SNPs, some of which may either alter the protein function or cause gene inactivation.
Two regions of the chromosome were found to have accumulated a large number of polymorphic sites unique within the *plantarum* strains. These regions mainly contain genes related to non-ribosomal synthesis of secondary metabolites. The high potential for production of various secondary metabolites, siderophores and other metabolites for members of the *B. amyloliquefaciens* subsp. *plantarum* group seems to be an important feature for the efficacy of providing beneficial effects to plants.

### 4.3.3 Carotenoid biosynthesis

UCMB5113 is dark orange pigment producing bacterium probably due to production of carotenoid. However, the intensity of pigment formation depends on composition of the medium, growth temperature, and light conditions (Paper IV). Many of the orange pigments (carotenoids) produced by bacteria exhibit biological roles, such as providing protection to photo-oxidative damage, reactive oxygen species, and influence membrane integrity (Olson & Krinsky, 1995; Britton 1995; Mishra et al., 2011). It has been observed that the pigment is only produced when the strain is grown under dark conditions. A gene probably encoding phytoene/squalene synthases, a key enzyme in the process of carotenoid biosynthesis, was found in the genomes of the three UCMB strains with high identity. The analysis revealed that only UCMB5113 encoded phytoene/squalene synthases and was highly expressed during pigmentation while the expression was significantly lower in the other two non-pigmenting strains. The production of the orange pigment by UCMB5113 probably controls environmental stress. Further studies are needed to completely characterize the molecule and its function in the cell.
5 Conclusions

The genomes of all three *B. amyloliquefaciens* isolates contain genes that are required for a free-living lifestyle. They have high capacity to produce several antibiotics and also exo-enzymes to improve nutrient acquisition in the rhizosphere. The potential to produce hormones and volatile compounds can also influence plant growth. These strains seem to have the capability to utilize a wide range of carbon and organic compounds available in the rhizosphere and in exchange improve nutrient uptake and stress handling capabilities of the host plant. The comparison of the three genomes indicates that they undergo similar biological processes involved in plant growth promotion and protection due to common genetic traits, which are found across the species. Though small genetic variation in genetic traits and regulatory apparatus exists between the genomes that may affect the underlying plant growth promotion and protection ability.
6 Future perspectives and implications

The results presented in this thesis provide a genomic overview of some plant-growth promoting *Bacillus* ecotypes. Elucidation of some of the putative genetic traits involved in antagonism against pathogens and plant-growth promotion mechanisms described in this thesis are the first step to foster more effective BCAs. Furthermore, the available information provides a good basis to investigate the evolution of plant growth promotion and biocontrol mechanisms in biocontrol microbial species by comparative genomics approach.

The successful application of PGPR in agriculture is challenging and, therefore, direly requires deep understanding of the hitherto unknown biological processes mediating cross talks and modes of action that make plant more tolerant to biotic and abiotic stress as well as trigger plant growth. Transcriptome profiling will be advantageous for the identification of functional traits involved in different plant-associated mechanisms and will help to provide an overview of interactions between microbes and plants. Several genes discussed in this thesis are putatively involved in physiological and metabolic processes but many hypothetical genes were also predicted without any known function. A combination of mutagenesis and proteomics studies will allow us to verify the putative genes and attribute biological functions to the unknown genes. Furthermore, the use of metagenomics approaches will be valuable to better understand the possible impact of biocontrol organisms on cohabiting microbial communities under natural conditions. There is very poor understanding about the role of host-associated bacteria in switching on and off the host genes. It will be valuable to elucidate role of bacterial association in genetic and epigenetic reprogramming of host plant genes.
References


Acknowledgements

This journey had never been possible without all colleagues, friends and family that have supported me during these years.

First of all I would like to thank my supervisor Erik Bongcam-Rudloff. Thank you for all the support and freedom to explore during this journey. Johan Meijer, thank you for your support and always taking time for me whenever I needed directions and guidance. Sarosh, thank you for all your help and patience, and interesting discussions. This would have been an impossible achievement without you.

Thank you to all colleagues, new and old. Thank you all my friends. It is impossible to mention everyone here but you know who you are. You all deserve thanks for providing well-needed distractions and friendship. Thank you for all discussions about science and politics as well as everything in life and making me feel welcome. I would like to thank Higher Education Commission (HEC) of Pakistan for providing scholarship during the period of research.

Lastly I would like to thank my family. My parents, my siblings, my wife and adorable daughter for always being there for me and always giving me their full support.